

CCL25-Eval 任务四系统报告： 现代汉语动词的叙实性语用推理机制

赵培翔¹✉ 李明珠¹ 梅立亚¹ 王芳¹ 高年鑫¹ 赵浪²

1 昆明学院人文学院；2 湖南师范大学新闻与传播学院

729591048,3514973615,1444959938,2789076610,1356289392,3331316907@qq.com

摘要

叙实性推理是一项与事件真实性判断密切相关的语义理解任务，主要关注语言表达中的事实性信息传递。本次测评任务基于沈家煊（2003）提出的“行、知、言”三域理论，对动词叙实性分类体系进行了进一步细分。这一改进不仅为汉语叙实性研究提供了更为精细的分析工具，还显著提升了大语言模型对“叙实性”语义的理解能力。测试结果表明，在不微调赛道上，我们团队在测试集上的最终正确率达到93.41%。

关键词： 叙实性推理；行、知、言；跨域动词

System Report for CCL25-Eval Task4: The pragmatic reasoning mechanism of factivity inference of modern Chinese verbs

Peixiang Zhao¹ ✉ Mingzhu Li¹ Liya Mei¹ Fang Wang¹ Nianxin Gao¹ Lang Zhao²

1 School of Humanities, Kunming University

2 School of Journalism and Communication, Hunan Normal University

729591048,3514973615,1444959938,2789076610,1356289392,3331316907@qq.com

Abstract

Factivity Inference (FI) is a semantic understanding task closely related to the judgment of event authenticity, mainly focusing on the transmission of factual information in language expressions. This evaluation task was based on the "Acting, Knowing and Uttering" three-domain theory proposed by Jiakuan Shen (2003), and further refined the classification system of verb factivity. This improvement not only provides a more refined analytical tool for the study of Chinese factivity but also significantly enhances the large language model's semantic understanding ability of "factivity". The test results show that in the non-fine-tuning track, our team achieved a final accuracy rate of 93.41% on the test set.

Keywords: Factivity Inference, Acting, Knowing and Uttering, Cross-domain Verbs

1 任务介绍

叙实性推理 (Factivity Inference, FI) 是自然语言处理领域的重要研究方向，专注于探讨语言使用者如何从动词性成分的语义预设中推导命题的真实性。作为自然语言处理的核心推理

©2025 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

范式之一，叙实性推理通过系统解析语言形式中编码的事实性标记（如叙实性谓词、反事实连接词、情态副词等），为机器构建了一套理解人类话语真实维度的结构化认知框架。这一机制不仅是文本蕴涵识别的逻辑基础——例如，从“他隐瞒检测结果”可以必然推导出“检测结果客观存在”；同时，它也是消除生成式AI幻觉的关键屏障，通过约束诸如“后悔未发生事件”这类语义矛盾的生成，确保语言模型的输出符合现实逻辑。在动态认知场景中，叙实性推理驱动信念修正系统穿透语言表层，精准区分“陈述事实”（如“实验失败”）与“表达态度”（如“谎称成功”）的本质差异，从而构建信息可信度的多维度评估矩阵。这种能力进一步向下游任务辐射：在信息检索中，能够识别“报告漏洞”与“假装发现漏洞”的本质区别；在情感分析中，能够辨析“遗憾失败”与“庆幸失败”的态度对立；在智能问答中，则能够规避“虚构历史事件”的推理陷阱。叙实性推理的独特作用使其成为提升语言理解和生成质量的重要基石，同时也为人工智能迈向更深层次的认知能力提供了关键支撑。

首届中文叙实性推理评测（FIE2025）重点考察大语言模型对汉语动词预设的理解能力。评测旨在系统评估不同大语言模型（LLMs）在中文叙实性推理任务中的表现差异，重点考察语境复杂度、提示词设计（如few-shot示例数量、思维链/CoT应用、提问句式调整）对推理效果的影响。基于官方测试集自主设计提示策略并选择适配模型，通过API交互探索最优解法（如融合知识增强与推理路径可视化），采用复合式提示工程（如“假设-验证”框架）突破现有模型对否定句、跨句指代等复杂语义的理解局限。评测开放模型选型与提问方式，旨在通过多样化测试手段挖掘LLMs的事实性推理潜力。

本次测评借鉴沈家煊(2003)的“行、知、言”三域理论对动词叙实性分类体系进一步细分，在不微调赛道上的正确率达到93.41%。实验结果表明，大模型在简单事实推理（单域动词）等基础叙实推理任务中表现优异，但在复杂语用环境下的推理能力仍存在显著局限，特别是在需要多层级语义整合（跨域动词）的任务中准确率下降明显。这些发现不仅为汉语叙实性研究提供更加精细的分析工具，而且进一步提升大语言模型对“叙实性”的语义理解能力，为智能体的语义理解能力优化提供了重要参考维度。

2 相关工作

本次测评任务是首届中文叙实性推理评测（FIE2025），目前尚无前人评测的任务来以供参考。国外研究以Paul Kiparsky & Carl Kiparsky(1970)提出的真值预设理论为根基，将动词划分为“叙实”与“非叙实”两类。Leech(1983)依据谓词对从属结构规定性质，将动词分为叙实动词、非叙实动词和反叙实动词三类。而Givón(2001)将叙实动词和非叙实动词都分为肯定的和否定的两类。在国内研究方面，袁毓林(2014)首次为汉语动词的叙实性提供了相对明确且系统的界定。在此基础上，国内(李新良, 2020; 李新良、袁毓林等, 2023; 陈振宇、甄成, 2017; 陈振宇、姜毅宁, 2019; 张新华, 2020)等学者进一步推动汉语动词叙实性的研究向更深层次发展。

主办方可能按传统研究方式，对测试集的1000条数据（人造语料400条，自然语料600条）进行测试，其基线结果如下：微调前人造语料正确率为53.74%，自然语料正确率为69.86%；微调后人造语料正确率为94.66%，自然语料正确率为88.93%。由此，微调前这种基于命题真值单向判定的二分法虽具形式逻辑的简洁性，却在汉语实践中遭遇三重困境：其一，忽略动词在不同语境中可能激活多重认知维度（如“承认”既可指向客观行为记录，也表示其主观立场）；其二，弱化言者交际意图对真值协商的动态干预（如外交辞令中“注意到”的叙实性弱化现象）；其三，难以解释新兴话语现象（如网络语境中“我觉得你死了”的非叙实用法）。

这些语言现象提示着，叙实性本质上是语言符号在行为实践、认知建构与言语交互的三维空间中的复合范畴。沈家煊(2003)将语言研究的“三域”理论框架系统划分为“行域、知域、言域”三个维度：行域指涉现实世界中的行为表征及事态呈现，关涉客观的行为过程与事件状态；知域聚焦认知主体的知觉系统与知识体系，体现言谈参与者的认知互动与信息处理机制；言域则对应于言语行为的施为功能，具体表征为指令、承诺、建议等具有交际意图的语用范畴。这三个维度共同构成动词叙实性存在的深层动因。袁毓林(2021)认为“假装”一词的使用有不同的认知层面和界域，会造成“假装”句语义识解的歧异性。受到两位学者启发，从“三域”角度对动词的叙实性进一步分类研究。（测评中涉及对动词的叙实性再分类详见附录1）

本次测评将测评语料进行分为单域动词与跨域动词两大类，单域动词分为正叙实和反叙实两类。正叙实动词又分为行域叙实动词、知域叙实动词、言域叙实动词三类；反叙实动词再分为一般反叙实动词与特殊反叙实动词两类。具体定义如下表所示：

动词叙实性再分类		定义	
单域动词	正叙实	行域叙实动词	行域正叙实动词，仅预设宾语小句内容为客观事实，客观描述现实的行为和行状，强调“事实不可辩驳性”。
		知域叙实动词	知域正叙实动词预设其宾语小句内容为真，但同时表达说话者对该事实的认知或情感，强调“主观事实性”。
		言域叙实动词	言域正叙实动词，通过言语行为本身直接预设宾语小句为事实，且该预设的成立依赖于说话者通过语言行为对事实的确认或宣告。
	反叙实	一般反叙实动词	典型反叙实动词指完全通过心理认知或语言宣称虚构非现实情境的一类动词，其宾语小句的命题在客观现实中为假，且无现实行为支撑。
		特殊反叙实动词	行域反叙实动词是反叙实动词的特殊子类，其宾语小句的命题仍预设为假，但主体在行域中执行了真实动作以虚构某种状态或目的。
跨域动词 (U \leftrightarrow T/F)		跨域动词是指在汉语中横跨多个认知域的动词类别。	

Table 1: 动词叙实性再分类以及定义¹²

3 测评所使用的软件与模型介绍

3.1 Trae软件

Trae是由字节跳动发布的AI原生编程工具，它与人工智能技术深度结合，具备智能问答、代码自动补全以及基于Agent的AI自动编程等多种功能。对于理论语言学研究者来说，这款软件为理论研究提供核心赋能：通过自然语言指令生成自动化代码，将“构建标注框架”等需求转化为Python脚本，打破语言学与编程的技术壁垒；基于智能体协作机制支持多角色分工与逻辑校验，通过上下文验证规避模型偏差风险，提升研究结果的可靠性与合规性。该工具以低代码化、开放模型接口和可信协作模式为核心，为语言学研究构建高效灵活的AI辅助科研路径。

3.2 “Deepseek-V3-0324”大语言模型

为了确保最终选用的模型能够在性能、准确率以及响应速度等方面满足测评的实际需求，我们不仅关注模型的基础生成能力，还特别重视其在中文复杂指令理解、逻辑推理能力以及输出稳定性的表现。在选择适用于本次测评任务的大模型过程中，本次测评以C-Eval与ZebraLogic评估指标作为参考标准，相关数据结果来源于相关技术报告（Table 2）。基于这些数据，最终决定选择“Deepseek”大语言模型。

模型/架构	C-Eval (中文评估)	ZebraLogic (逻辑推理)
GPT-4o-2024-11-20	75.5	27.4
DeepSeek-V3	86.5	42.1
Qwen2.5-72B-Instruct	84.7	26.6
LLaMA-4-Maverick	83.5	40.0
Qwen3-235B-A22B	86.1	37.7

Table 2: LLMs (Non-thinking) 在C-Eval与ZebraLogic指标上的得分对比

注：表格数据来自Qwen 3的技术报告(Qwen Team, 2025)

¹鉴于特殊反叙实动词（假装）具有争议，主办方未将其纳入正确率的计算范围。这类动词虽被划分到特殊反叙实动词，但与句子语境相结合后呈现出“多声性”特征(袁毓林, 2021)，正处于单域动词与跨域动词的交界地带。有关这类动词研究可参见将在第十三届现代汉语语法国际研讨会上宣读的论文——“汉语‘假装’类的三维认知模型研究”。

²当“跨域动词”出现时，命题的真伪由评价主体的情感、态度或社会角色决定，需通过互动协商（如支持、反驳）形成立场对齐，此时这类动词偏向非叙实动词，则宾语小句真值为“U”；当命题的真实性独立于主观立场，此时这类动词偏向叙实动词（包含反叙实动词），则宾语小句真值为“T/F”。

最初尝试使用了深度推理模式下的“Deepseek-R1”模型，期望通过增强推理能力来提升模型处理复杂语言学问题的效果。然而，在实际应用中却发现，尽管该模式在一定程度上增强了模型的逻辑推理能力，但也带来了诸如“幻觉现象”频发、对用户指令理解偏差等问题。这些问题在语言学研究尤为敏感，因为它们可能导致分析结果的失真或误导性结论的产生，严重影响研究的严谨性和可信度。面对这些问题，我们将目光转向了Trae平台，并在其中尝试“Deepseek-V3-0324”这一版本的模型。该模型在未开启深度推理的情况下，依然展现出了极高的响应速度和出色的准确性。更重要的是，它表现出更强的指令遵循能力和较低的幻觉发生率，能够更稳定地完成指定的语言处理任务，从而提升了整个研究流程的质量与效率。

4 测评方法介绍

本次测评采用“Trae CN”平台中搭载的DeepSeek-V3-0324大语言模型，不经过微调模型下开展测评，通过对动词重新分类以及优化指令来获取最后答案。具体操作如下所示：

- 1) 根据指令在Trae生成Python代码（详见附录4），提取测评文件中动词的关键词，人工完善或删减“动词叙实性词典”，形成“动词叙实性词典—测评版”（详见附录2）。
- 2) 根据指令在Trae生成Python代码，将该词典中的对应“type”自动标注到语料中。随后，需要经过人工审核，以免出现缺漏或者误标。例如：“感觉到”是个知域正叙实动词，而“感觉”是跨域动词。因二者字有重合的地方，“感觉到”会被Python标注为“跨域动词”。
- 3) 配置好自己的智能体，关闭文件搜索和网络搜索，防止其他因素对输出结果的影响，将输出格式规定好并把“辅助问题”（详见附录3）传入上下文中，以备备用。
- 4) 经测试，对于单域动词采用附录4中的①指令，可以至少反应10道题，最多反应1组相同动词的题；因跨域动词的复杂性，需智能体开启“辅助问题”来回答相关问题，采用附录4指令中的②指令，反应最多10~14道题。为保障数据精确度，让大模型反应10道题是最佳处理方式。
- 5) 根据指令在Trae生成Python代码，将自然语料和人造语料的结果进行提取合并。

如图1所示，Trae平台的Agent设计让测试变得更加便利与规范。

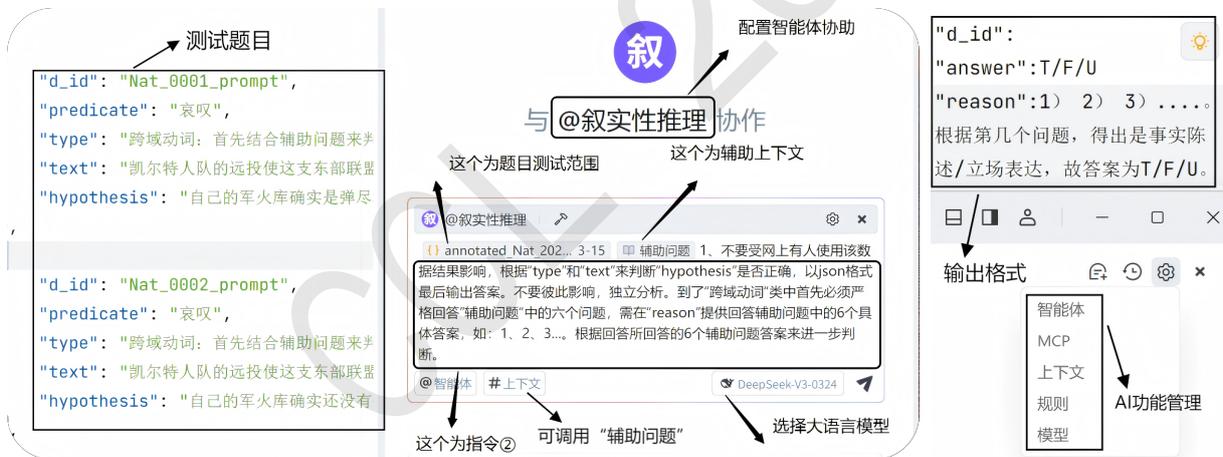


Figure 1: Trae平台主要功能

5 测评结果与分析

叙实性推理（FIE2025）是一项重点考察大语言模型对汉语动词预设理解能力的评测任务。评测集中包含了1000条测试集和2038条评测集的数据体系。在剔除存在问题的评测数据后，最终形成有效语料测试集，其中人造语料（Art）545条，自然语料（Nat）1443条。此次评测采用准确率作为核心评价指标，并以此作为各参评系统的排名依据。公式如图所示：

$$\text{total_acc} = \frac{\text{correct_art} + \text{correct_nat}}{\text{total_art} + \text{total_nat}}$$

参赛队伍数量较多（共计74支），本文提出的方法在不微调赛道中取得了第四名的成绩。通过分析提交的三次结果（Table 3），可以发现：

测评	correct_art	correct_nat	total_acc
5.7测评	0.9534	0.9088	0.9210
5.11测评	0.9624	0.9020	0.9185
5.12测评	0.9761	0.9182	0.9341

Table 3: 测评结果对比

在人造语料方面，正确率逐步提升。5.7测评采用上下文嵌入式指令框架达到95.34%准确率；5.11测评通过Python脚本提取某些动词题目，有效减少答案输出与原因解释的偏差（如：答案是U，但在reason中根据相关证据是得出答案是T，前后不一致），准确率提升至96.24%；5.12测评引入指令驱动式生成范式，要求分步骤输出语义角色标注，成功将准确率推升至97.61%，充分验证了结构化指令对模型推理能力的增强效应。

在自然语料方面，正确率因测评中调整测评方法而产生一定波动。5.11测评尝试将“跨域动词”体系再进一步细分为“一般跨域动词”与“特殊跨域动词”两类，取消原有“立场表达”与“事实陈述”的宾语小句区分。由于新增子类没有足够样本和时间去验证，导致正确率反向波动0.68%，跌至90.2%。因结果表现不佳，只能对此方法进行放弃，在第三次测评中做保守处理。在5.12测评中采取混合策略：一方面恢复5.7测评的上下文处理框架，另一方面将辅助问题转化为显式指令输入，而非延续5.7测评中的隐式上下文嵌入模式。该策略使自然语料准确率回升至91.82%。经过后来分析后发现，跨域动词错误率仍较高，揭示出现有模型在语言迁移现象处理中的深层局限，也为下一步的研究提供较好的研究对象。

6 结语

本次测评在结合前人研究与大模型技术基础上，从“三域”视角切入，对动词的叙实性进行重新分类。面对一些较为简单的动词（一般是单域动词），利用动词的认知域可以对其宾语小句真值做出简要判断。面对较为复杂的“跨域动词”，就看宾语小句是客观表述还是立场表达。这种分类，对于动词的叙实性分类有了进一步发展，更重要的是为LLM提供混合型推理模型，简单的快速处理，复杂的简单处理，以此来提高大模型对汉语动词叙实性的理解能力。

经过本次测评发现，大语言模型在处理单域动词的叙实性语义理解任务中展现出显著优势。模型能够对于单域动词的叙实性推理的正确率较高。这种优异表现主要源于对动词在叙实性上面得以较为精确的分类，使得模型能够正确理解动词背后的语义蕴含。然而，当面临跨域动词叙实性理解时，模型的正确率较低。因此，针对跨域动词的叙实性语义理解的进一步研究与改进，将是今后需要研究的重点。例如：可尝试将“跨域动词”进一步细化分类为一般和特殊两类；或者借鉴在第十三届现代汉语语法国际研讨会上即将提出的关于汉语“假装”的三维认知模型，能否将其迁移应用于“跨域动词”的研究之中；抑或是把“跨域动词”与Talmy(2000)所提出的力动态模式（Force Dynamics）相结合来进一步深化动词叙实性研究等。这些都为后续研究提供了值得探索的路径。

参考文献

- Geoffrey Leech. 1983. *Semantics: The Study of Meaning* (2nd ed.). Penguin.
- John W. Du Bois. 2007. *The stance triangle*. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). John Benjamins.
- Leonard Talmy. 2000. *Toward a Cognitive Semantics: Concept Structuring Systems*. The MIT Press.
- Paul Kiparsky & Carol Kiparsky. 1970. *Fact*. In M. Bierwisch & K. Heidolph (Eds.), *Progress in linguistics* (pp. 143–147). Mouton.
- Qwen Team. 2025. *Qwen3 Technical Report*. arXiv:2505.09388.
- T. Givón. 2001. *Syntax: An Introduction. Volume II*. John Benjamins Publishing Company.

- 陈振宇. 2017. 汉语的指称与命题. 上海人民出版社, 上海.
- 陈振宇,甄成. 2017. 叙实性的本质——词汇语义还是修辞语用. 当代修辞学, (1):9-23.
- 陈振宇,姜毅宁. 2019. 反预期与事实性——以“合理性”语句为例. 中国语文, (3):296-310,382-383.
- 郭锐. 1997. 过程和非过程——汉语谓词性成分的两种外在时间类型. 中国语文, (3):162-175.
- 李新良. 2020. 现代汉语动词的叙实性研究. 北京大学出版社, 北京.
- 李新良,袁毓林. 2016. 反叙实动词宾语真假的语法条件及其概念动因. 当代语言学, 18(2):194-215.
- 李新良,袁毓林等. 2023. 叙实性与事实性理论及其运用. 外语教学与研究出版社, 北京.
- 沈家煊. 1999. 不对称和标记论. 江西教育出版社, 南昌.
- 沈家煊. 2003. 复句三域“行、知、言”. 中国语文, (3):195-204,287.
- 袁毓林. 2014. 隐性否定动词的叙实性和极项允准功能. 语言科学, 13(6):575-586.
- 袁毓林. 2015. 汉语反事实表达及其思维特点. 中国社会科学, (8):126-144,207.
- 袁毓林. 2020. 叙实性和事实性: 语言推理的两种导航机制. 语文研究, (1):1-9.
- 袁毓林. 2021. 从语言的“多声性”看“假装”句的解读歧异. 语言战略研究, 6(5):77-90.
- 张新华. 2020. 汉语叙实谓词研究. 复旦大学出版社, 上海.

附录

附录1: 测评中所涉及到的动词分类表

叙实性动词分类		行域动词	知域动词	言域动词
单域动词	正叙实 (T)	①发现、目睹 ②听出、看出、闻出、反映出 ③看见、瞥见、听见、望见、遇见、瞧见、碰见、闻见 ④碰到、看到、闻到、听到	①猜到、感觉到、觉察到、领悟到、预见到、注意到、认识到、意识到; 觉察出、觉出、暴露出; 披露、揭露; 得知、获悉、记得、明白、清楚、忘记、知道 ②后悔、诧异、高兴、庆幸、羡慕	表明、证明、证实、承认
	反叙实 (F)	假装、装作	幻想、想象	吹嘘、谎称、妄称、污蔑、诬陷
跨域动词 (U ⇔ T/F)		猜、想、说; 猜测、感觉、觉得、觉着、怀疑、以为、认为、相信、推算、估计、抱怨、埋怨、批评、数落、哀叹、埋怨、感叹、声称、声明、声言、听说、扬言、重申		

附录2: 动词叙实性词典—测评版

1、“predicate”：“发现、目睹、听出、看出、反映出、看见、瞥见、听见、望见、遇见、瞧见、碰见、碰到、看到、闻到、听到”

“type”：“行域正叙实动词：行域正叙实动词仅预设宾语小句内容为客观事实，客观描述现实的行为和行状，强调‘事实不可辩驳性’。‘否定词+行域正叙实动词’仅表示未感知，其宾语小句真值不定。”

2、“predicate”：“猜到、想到、认识到、感觉到、觉察到、领悟到、预见到、注意到、意识到; 觉察出、觉出、暴露出、揭露、披露、忘记、得知、获悉、记得、明白、清楚、知道、后悔、诧异、高兴、庆幸、羡慕”

“type”：“知域正叙实动词：预设其宾语小句内容为真，但同时表达说话者对该事实的认知或情感，强调‘主观事实性’。‘否定词+知域正叙实动词’下，其宾语小句也预设为真。”

3、“predicate”：“表明、证明、证实、承认”

“type”：“表示积极正向的言域叙实动词，预设宾语小句内容为真。‘否定词+表示积极正向的言域叙实动词’下，其宾语小句真值不定。”

4、“predicate”：“幻想、想象、吹嘘、谎称、妄称、污蔑、诬陷”

“type”：“一般反叙实动词：指完全通过心理认知或语言宣称虚构非现实情境的一类动词，其宾语小句的命题在客观现实中为假，且无现实行为支撑。”

5、“predicate”：“假装、装作”

“type”：“特殊反叙实动词：行域反叙实动词的宾语小句的命题仍预设假，但主体在行域中执行了真实动作以虚构某种状态或目的。”

6、“predicate”：“猜、想、说、声称、声明、声言、扬言、重申、感觉、觉得、觉着、怀疑、以为、认为、猜测、相信、推算、估计、抱怨、批评、数落、哀叹、埋怨、感叹”

“type”：“跨域动词：首先结合辅助问题来判断事实陈述与立场表达。如果宾语小句为立场表达，则该动词更偏向非叙实动词，其宾语小句真值为U；如果宾语小句为事实陈述，则该动词更偏向叙实动词（包括反叙实动词），其宾语小句真值为T/F。‘否定词+跨域动词’不会影响宾语小句的真值，由宾语小句是立场表达还是事实陈述来决定。”

附录3：辅助问题以及回答¹

(1) 上下文是否提供语境支撑（如数据、亲历描述、权威结论、相关证据支撑等），使原本立场表达的宾语小句转为事实？

例：原句“他说地球是平的”为立场表达，但若语境补充“根据卫星观测数据，地球是球形”→转为事实陈述

(2) 宾语小句的时态与体态是什么？

①时态：发生在一般过去时或一般现在时→强制为事实陈述；

发生在未来→强制为立场表达

例：“他昨天赢了比赛。/他没去学校。/他经常上课睡觉。”→事实陈述

“他明天会赢比赛”→立场表达

②体态：动作表示完成时→事实陈述；动作未完成或表示进行时→立场表达

例：“实验已证明结论”（事实陈述）vs“实验正在验证假设”（立场表达）

(3) 句式结构判断：句子整体是否为疑问句（难怪）或条件句（如果.../除此）？

若是，是否隐含对命题真值的肯定/否定（如“如果下雨，比赛会取消吗？”隐含“下雨可能导致取消”）→事实陈述

若仅表达可能性或不确定性（如“如果下雨，你会来吗？”）→立场表达

(4) “predicate”中的动词之前是否有强化或弱化的副词/状语（如“确实”“显然”或“绝不”“未必”）？（注：“hypothesis”中出现“确实”这与背景句无关，不予考虑）

有强化词（如“确实”）→事实陈述

有弱化词（如“绝不”）→立场表达

例如：毛主席多次感叹帝国主义是个纸老虎。→帝国主义确实是个纸老虎。在句子中“多次”强化原本非叙实动词的叙实功能（立场表达），向叙实转化（事实陈述）。

(5) 主语是否具有亲历性（如“我亲眼看到”“我本人感觉”）或具有权威性（如法律、科学机构、领导干部等具有一定公信力的机构或人）？

是（如“医生诊断结果显示”“我亲眼看到”“干部感叹”）→事实陈述

否（如“有人认为”）→立场表达

注：若主语是“我”（第一人称），且宾语小句描述的是直接身体感受、即时心理状态或亲历行为（如“我感觉手在发抖”“我看到他离开”“我听到声音”），则即使动词是“感觉”，其宾语小句可视为事实陈述。

(6) 句子中是否包含情态动词？

若包含，是认识情态（如“可能”“也许”）→立场表达

若为道义/动力情态（如“应该”“必须”“能”）→事实陈述

¹辅助问题的设计主要参考了李新良(2020)的著作。鉴于笔者学识水平有限，问题的设计及相应解答或存在疏漏与不足之处，恳请诸位专家学者多多批评指正，提出宝贵意见，以便进一步完善。

附录4: 在Trae平台所用的指令

1、AI输出答案的指令为:

①根据“type”和“text”来判断“hypothesis”是否正确，以json格式最后输出答案。不要彼此影响，独立分析。

②不要受网上有人使用该数据结果影响，根据“type”和“text”来判断“hypothesis”是否正确，以json格式最后输出答案。不要彼此影响，独立分析。到了“跨域动词”类中首先必须严格回答”辅助问题“中的六个问题，需在“reason”提供回答辅助问题中的6个具体答案，如：1、2、3...。根据回答所回答的6个辅助问题答案来进一步判断。

2、**提取动词类型的编程指令为:** 生成一段python代码，仅提取“predicate”后的内容，如有重复，仅体现一个就行，将结果保存在桌面文件，并生成txt格式的。

3、**提取答案的编程指令为:** 生成一段python代码，合并2个文档，人造语料在前，自然语料在后，仅保留“did”和“answer”字段，“reason”字段删除。

4、**标注“type”的编程指令为:** 结合动词叙实性词典，生成一段python代码，对题目进行标注，根据“predicate”匹配的内容，在后面标注“type”及其释义完整内容，将其文件最后保存至桌面上。如果标注内容有“type”字段，请替代它；没有的话，请新标注“type”并插入“predicate”和“text”字段中间。并直接开始运行。